International Journal of Engineering & Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-3444

International Journal of Engineering and Technology Volume 2 No. 7, July, 2012

Security Threats Prediction on Local Area Network Using Regression Model

[1]Alese B.K. , [2]Osisanwo F.Y [3]Fasoranbaku O., [4]Adetunmbi A.O.
[1,3,4] Federal University of Technology, Akure
[2]Adeyemi College of Education, Ondo

ABSTRACT

In this dispensation, security of data and information has become a great threat to organisations involved in the use of networking. This is as a result of the increase in the rate of spoofing, hacking, eavesdroping, password breaking among others.

Vulnerability scanners now play a major role in identifying vulnerabilities or threats that exist across an organisations Local Area Network. When an organisation becomes aware of vulnerabilities that can be exploited on its LAN, the organisation becomes prepared on how to maintain and prevent further exploitation by malicious users. This work focuses on predicting vulnerablities of Local Area Network (LAN) that is likely to occur in nearest future with the use of Econometric method(Regression Model) on previous or historical data gathered from the Local Area Network (LAN). Data from two case studies of a previous research were sampled. Finally, a comparative analysis of the results from the statistical model (Econometric model) employed was carried out with that of the previous statistical model (Box and Jenkins model).

Keywords: *LAN, hacking, eavesdropping, vulnerabilities, threats*

1. INTRODUCTION

Local Area Network (LAN) supplies networking capability to a group of computers in close proximity to each other. This is useful for sharing resources like files, printers, games or other applications. A LAN in turn often connects to other LANs, and to the Internet or other WAN (Bradley, 2008). Local area networks have become a major tool to many organizations in meeting data processing and data communication needs, since it is a user network whereby data is sent at high rates between people located relatively close to each other. Prior to the use of LANs most processing and communication were centralized; even the information and control of that information were centralized as well. Now LANs logically and physically extend data, processing and communications facilities across the organization. (FIPS PUB191, 1994, Martin 1989, Matt 1991 and Michael, 2008).

As LAN technology expands through organizations, the organizations security becomes more porous, they are exposed to various threats than what it used to be. *Vulnerabilities* are weaknesses in a LAN that can be exploited by a threat therefore breaking the network securities and posing more threats or danger to the security of the LAN, this results to security threats.

Mark (1997) described *Security threats* as those threats that break through the security mechanism of an organization’s network due to the vulnerability of the network. Security threats can be categorized into Physical security threats and Network security threats. Where the physical threats are seen as more than just network media and servers; they also talk about clients and client environment. Network security threats are those threats associated with networked systems, the threats are launched on the resources of the network, i.e. the systems and other resources shared including files and information shared on the network.

Vulnerability scanners are used to identify those vulnerabilities. These vulnerability scanners are able to identify the potential flaws in the system. At present there are several vulnerability scanners available as both commercial and freeware, like Nessus, SAINT, ISS (Backley, 1989, Rich 2002).

2. VULNERABILITY PREDICTION

Somak and Gosh (2007) defined vulnerability predictions as an attempt to identify potential vulnerable areas on hosts across a network and the extent to which such areas on host will be vulnerable over a specific period of time in the near future. The aim of Vulnerability Prediction is to predict the number of known vulnerabilities that could occur and may be give a ranking of these vulnerabilities depending on their severity and impact which could lead to an efficient risk management procedure.

Statistical models are used for predictions purposes; the models used for predictions do have two main components according to Somak and Gosh (2007). The *data collector* which collects data through vulnerability scanners and the *data analyzer* which analyses data collected through forecasting techniques.

2.1 Forecasting Techniques

Forecasting is a prediction of what will occur in the future, and it is an uncertain process. A method for translating past experience into estimates of the future. Forecasting is much more than projecting a series mechanically into future. It involves making assumptions about the future course of an activity. Assumptions are made regarding the future on the basis of observations of the past. Forecasting technique is used in various facet of life to predict and plan for the future in order to survive and grow. Organizations, government, even nations plan for the future by forecasting. Plans for the future cannot be made without forecasting events and the relationship they will have.

Organizations and their executives have recognized the importance of forecasting as the basis of rational decisions and actions concerning the future. (Ya-lun, 1975). Statistical approach or methods are used for the forecasting processes there are various models that can be used for forecasting statistically this leads to forecasting techniques. One statistical model whose major aim is forecasting is *Time series analysis (Bowerman and O'Connel, 1987)*.

Also Ya-lun (1975) in the book statistical analysis states that numerous forecasting techniques with varying degrees of complexity have been devised during the past few decades. Most of these fall into one of three broad categories: the naïve method, barometric method, and the analytical method.

2.2 Quantitative Forecasting Methods

The quantitaive methods according to Arsham (1994) is also seen as a method that uses time series for forecasting. Time series forecasting methods are based on analysis of historical data (time series: a set of observations measured at successive times or over successive periods). They make the assumption that past patterns in data can be used to forecast future data points.

Bruce (1993) defined time series as a set of data collected at successive points in time or over successive periods of time. Also Somak and Gosh (2007) defined time series as a series where data is taken at successive times spaced apart at uniform time intervals. Meanwhile, Bruce and Richard (1993) stated that; the two most widely used methods of forecasting are:

a. Box-Jenkins autoregressive integrated moving average (ARIMA)

b. Econometric methods.

3. ECONOMETRIC MODEL FOR VULNERABILITY PREDICTION

Econometrics is concerned with the tasks of developing and applying quantitative or statistical methods to the study and elucidation of economic principles. Econometrics combines economic theory with statistics to analyze and test economic relationships. Works in econometrics focused on time-series data. The Econometric methods develop forecasts of a time series using one or more related time series and possibly past values of the time series. This approach involves developing a ***regression model*** in which the time series is forecast as the dependent variable the related time series as well as the past values of the time series are the independent or prediction variables.

In statistics, **regression analysis** refers to techniques for the modeling and analysis of numerical data consisting of values of a dependent variable (also called a response variable) and of one or more independent variables (also known as explanatory variables or predictors). Three stages are involved in using regression for data analysis these are Model Identification, Parameter or Data and Estimation and Analysis or Prediction (David, 19io 2001 and Paul 1998).

3.1 Model Identification

Regression attempts to model the relationship between two variables by fitting a linear equation to observed data. Before attempting to fit a linear model to observed data, it should first be determine the type of relationship that is between the variables of interest, whether or not it is linear relationship. A scatterplot or scatter diagram can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., the scatterplot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model, until the relationship is transformed to be so.

A linear regression equation with one independent variable represents a straight line when the predicted value (i.e. the dependant variable from the regression equation) is plotted against the independent variable: this is called a simple linear regression. A linear regression line has an equation of the form

Y= a + bX or this can be expressed as (1)

$$Y = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2)$$

where ***X*** is the explanatory variable and ***Y*** is the dependent variable, The slope of the line is ***b***, and ***a*** is the intercept. Or can also be explained thus: There is one independent variable: X_i, and two parameters, β_0 and β_1:

The multiple linear regression model takes the following form

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i \\ i = 1, \ldots, n \quad (3)$$

Non linear functions can be transformed or reduced to linear relationship by appropriate tranformation of variables. Then standard linear regression can be performed but with caution. The linearization can be done either by taking the natural logarithm value and differencing between two consecutive values of the resulting values or by finding the second, third or even fourth power of the independent variable and using the original variable and the once whose power has been raised as the independent variable.

3.1.1 Linearisation by Taking the Natural Logarithms

Here to linearize the non linear data so as to have a linear data for the equation, the natural logarithm of the Vulnerability counts were taken to get log(Y_t) then the differences of the consecutive values of the Log(Y_t) were taken to give values for new set of data labelled or refered to as Y_t, the differences of the consecutive values of Y_t were taken to give ΔYt, where, $\Delta Yt = Y_t - Y_{t-1}$. Also the differences of the consecutive values of ΔYt were taken and the set of resulting data were labelled $\Delta^2 Yt$, where $\Delta^2 Yt = \Delta(\Delta Yt) = \Delta(Y_t - Y_{t-1})$ or $\Delta Y_t - \Delta Y_{t-1}$ .This differencing is done continuously up to $\Delta^n Yt$, where $\Delta^n Yt = \Delta(\Delta^{n-1} Yt)$. The value of n in $\Delta^n Yt$ is determined by the R square (correlation coefficient). This is because R square measures the goodness of fit for the model. It determines if the equation of line is accurate.

R^2 equals the square of the correlation coefficient between the observed and modelled (predicted) data values. It is sometimes calculated as the square of the correlation coefficient between the original and modelled data values. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression line approximates the real data points. An R^2 of 1.0 indicates that the regression line perfectly fits the data.

The R^2 is calculated using the expression below

$$R(X,Y) = [\mathrm{cov}(X,Y)]/[stdDev(X) * stdDev(Y)] \quad (4)$$

$$R = \frac{\mathrm{E}[x - \bar{x}][y - \bar{y}]}{\sigma x \sigma y} \quad (5)$$

Here the R square value was determined with the use of a statistical package Microsoft Excel.

3.2 Parameter Estimation

This research is based on Multiple Regression which is expressed below

$$Y_i = \beta_0 + \beta_1 X_{il} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i,\ i = 1,\ldots,n \quad (6)$$

Typically, from this expression the observed values, or data, consist of *n* values

$$(Y_i, X_{i1,\ldots,} X_{ip}),\ i = 1,\ldots,n$$

and then there are up to p + 1 no of parameters to be estimated: $\beta_0, \ldots, \beta_p$

So as to estimate this parameter matrix notation has to be used

Where Y is a column vector that includes

$$Y = \begin{bmatrix} Y1 \\ Y2 \\ . \\ . \\ . \\ Yn \end{bmatrix} \quad (8)$$

Then X the observed value of the regressors

$$X = \begin{bmatrix} 1 & X_{11} & X_{1P} \\ 1 & X_{21} & X_{2P} \\ 1 & X_{31} & X_{3P}. \\ . & \ldots & \ldots \\ . & \ldots & \ldots \\ 1 & X_{n1} & X_{nP} \end{bmatrix} \quad (9)$$

The first column in X is a constant column, which represent the intercept β_0 since it does not vary across the observation. The first objective of regression analysis is to best-fit the data by estimating the parameters of the model

$$\beta = (X^T X)^{-1} . (X^T Y) \quad (10)$$

This implies that the parameter estimators are linear combinations of the dependent variable. Prediction is then carried out with the estimated parameters

4. SAMPLE DATA 1

In this case study, the data has been collected from a previous research paper, with total number of *network and system information gathering* vulnerabilities detected in their network for the last 10 days. The collected data is shown in Table 1 below and the scatter plot or diagram is shown on figure 1 below. From the scatter diagram it is observed that the relationship between the observed data is not linear

Table 1: Collected Data of network and system information gathering Vulnerability

Time (Days)	Vulnerability Count(VC)
1	19
2	21
3	20
4	71
5	69
6	78
7	24
8	25
9	75
10	79

4.1 Model Identification

From the scatter diagram, It was observed that the relatioship is not linear. Then the variable was transformed to linear using one of the methods of linearizing a non linear time series data which is transformation by taking the Natural Logarithms. The natural logarithm of the Vulnerability counts were taken to get log(Y_t) then the differences of the consecutive values of the Log(Y_t) were taken to give values for new set of data labelled or refered to as Y_t, the differences of the consecutive values of Y_t were taken to give ΔYt, where, $\Delta Yt = Y_t - Y_{t-1}$

Also the differences of the consecutive values of ΔYt were taken and the set of resulting data were labelled Δ^2Yt, where $\Delta^2Yt = \Delta(\Delta Yt) = \Delta(Y_t - Y_{t-1})$ or $\Delta Y_t - \Delta Y_{t-1}$

This is shown in table 2 below

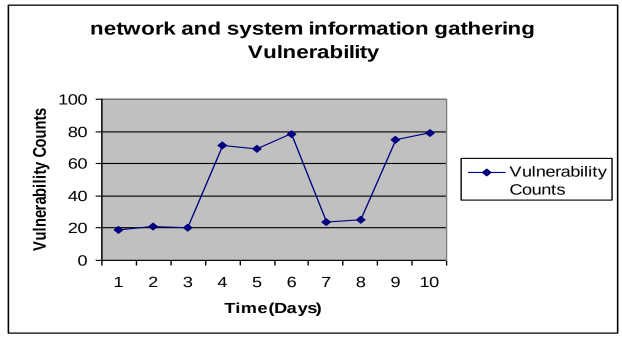

Figure 1 Scatter Diagram of original Time Series

From table 2 above, data set Y_t is taken as the dependent variable, while ΔYt, Δ^2Yt, and Δ^3Yt represent X_1, X_2, and X_3 respectively and they are the independent variable. The model regression equation for the data is

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Table 2 Linearization of variable Y

Y	Log(Y)	Y_t	ΔYt	Δ^2Yt	Δ^3Yt
19	1.278754				
21	1.322219	0.043466			
20	1.30103	-0.02119	-0.06465		
71	1.851258	0.550228	0.571418	0.636073	
69	1.838849	-0.01241	-0.56264	-1.13406	-1.77013
78	1.892095	0.053246	0.065655	0.628292	1.762348
24	1.380211	-0.51188	-0.56513	-0.63078	-1.25908
25	1.39794	0.017729	0.529612	1.094741	1.725525
75	1.875061	0.477121	0.459392	-0.07022	-1.16496

4.2 Parameter Estimation

To estimate the constant β_0 and the unknown parameters β_1, β_2, β_3 in the model equation $Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. Y_t is regress on X_1, X_2, and X_3. This can be done with matrix. The use of matrix is illustrated as follows as the data is converted to matrix formX will be represented in a 5 by 4 matrix where column 1 represent X_0, column 2 represent X_1, column 3 represent X_2, and column 4 represent X_3

$$X = \begin{bmatrix} 1 & -0.56264 & -1.13406 & -1.77013 \\ 1 & 0.065655 & 0.628292 & 1.762348 \\ 1 & -0.56513 & -0.63078 & -1.25908 \\ 1 & 0.529612 & 1.094741 & 1.725525 \\ 1 & 0.459392 & -0.07022 & -1.16496 \end{bmatrix} \quad (10)$$

while Y will be a vector matrix a 5 by 1 matix

$$Y = \begin{bmatrix} -0.01241 \\ 0.053246 \\ -0.51188 \\ 0.017729 \\ 0.477121 \end{bmatrix} \quad (2)$$

as well as the constant term and unknown parameter which is to be estimated

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad (11)$$

$$\beta = \left(X^1 X\right)^{-1}\left(X^1 Y\right) \quad (12)$$

Statistical package like SPSS and Ms-Excel could be use to Regress Yt on X_1, X_2, and X_3. In this case, Ms – Excel was used and from the output the values of the constant term β_0 is 0.06035 while β_1 is 1.762893, β_2 is -1.56163, and is β_3 is 0.458747.

4.3 Prediction

From the parameter estimates new set of values are predicted for Y which is equivalent to the differened values of the consecutive logarithm of Y. Suming the predicted value with the differenced value gives the Log of Y

Log Y_t = Predicted Y_t + Log Y_{t-1}

$Y_t = 10$ ^ $^{(\text{Predicted Yt + Log Yt-1})}$

Substituting the data into

Log Y_{10} = Predicted Y_{10} + Log Y_9

Log Y_{10} = 1.875061

$Y_{10} = 10$ ^$^{1.875061}$

Y_{10} = 75

The predicted value for the 10th day is 75. The measured value is 79.

4.4 Sample Data 2

In this case study, the data has been collected from a previous research paper as case study 1, the collected data in the case study has Vulnerability Counts with unauthorised access for 48 weeks and the 49th week is predicted. The same procedure as in case study one is used here but because larger set of data is involved here unlike in the case study 1 that has just few data set, the regression of variable is done twice, the predicted value is equally regress to get a more accurated prediction.

4.4.1 Model Identification

The collected data in case study two is shown above in *table.3* and the scatter diagram is equally on *figure 2.* From the scatter diagram, it is observed that the relationship between the variable is non linear, it has to be linearised by taking the natural logarithm of the VC and finding the difference between the consecutive values to give a new Y_t just as done in case study 1 various values for Y_t , ΔYt, $\Delta^2 Yt$, $\Delta^3 Yt$, $\Delta^4 Yt$ and so on. Depending on the value of the correlation coefficient (R Square). This is shown in table 4 below

From the table 4: Yt = log Y_2 – log Y_1

$$\Delta Yt = Y_t - Y_{t-1} \quad (13)$$

$$\Delta^2 Yt = \Delta(\Delta Yt) = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} \quad (14)$$

$$\Delta^n Yt = \Delta (\Delta^{n-1} Yt) \quad (15)$$

Table 3 collected Data of remotely exploitable vulnerabilities with unauthorised access

Time (weeks)	VC	Time (weeks)	VC	Time (weeks)	VC	Time (weeks)	VC
1	27	13	15	25	5	37	15
2	9	14	13	26	9	38	26
3	15	15	12	27	6	39	9
4	22	16	18	28	12	40	13
5	11	17	14	29	15	41	24
6	14	18	12	30	16	42	14
7	13	19	16	31	21	43	36
8	16	20	26	32	12	44	24
9	15	21	11	33	12	45	20
10	15	22	14	34	17	46	14
11	11	23	20	35	9	47	22
12	6	24	24	36	21	48	20

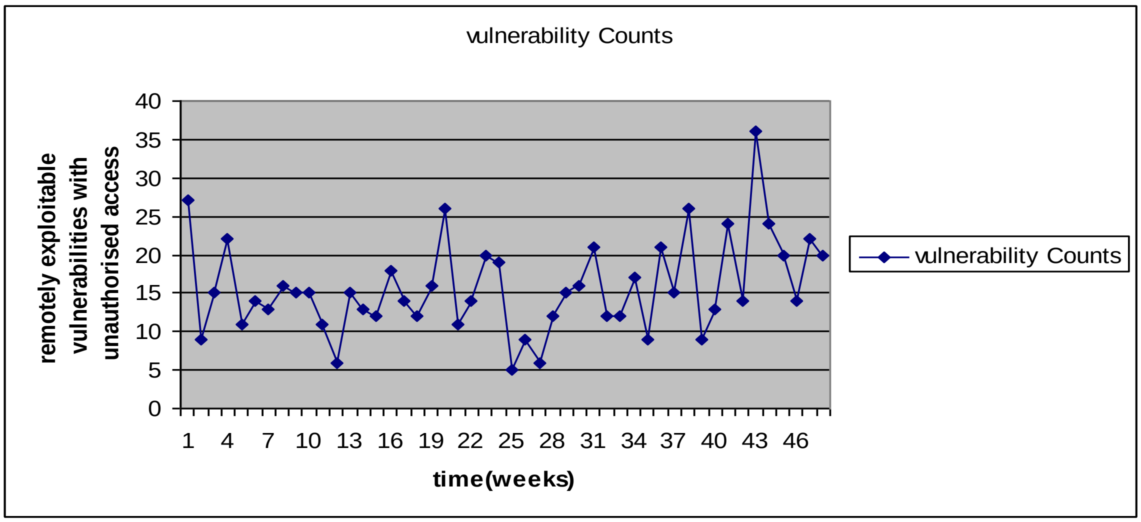

Figure 2 Scatter Diagram of Original Time Series Data

Table 4: linearising Y and generating Y_t and other variable values

Y	Log(Y)	Y_t	ΔYt	$\Delta^2 Yt$	$\Delta^3 Yt$	$\Delta^4 Yt$	$\Delta^5 Yt$	$\Delta^6 Yt$
27	1.431364							
9	0.954243	-0.47712						
15	1.176091	0.221849	0.69897					
22	1.342423	0.166331	-0.05552	-0.75449				
11	1.041393	-0.30103	-0.46736	-0.41184	0.342643			
14	1.146128	0.104735	0.405765	0.873127	1.284971	0.942328		
13	1.113943	-0.03218	-0.13692	-0.54269	-1.41581	-2.70078	-3.64311	
16	1.20412	0.090177	0.122361	0.259281	0.801967	2.217779	4.918562	8.561672
15	1.176091	-0.02803	-0.11821	-0.24057	-0.49985	-1.30181	-3.51959	-8.43816
15	1.176091	0	0.028029	0.146234	0.386801	0.886649	2.188464	5.708057
11	1.041393	-0.1347	-0.1347	-0.16273	-0.30896	-0.69576	-1.58241	-3.77087
6	0.778151	-0.26324	-0.12854	0.006156	0.168883	0.477844	1.173607	2.756017
15	1.176091	0.39794	0.661181	0.789724	0.783569	0.614686	0.136841	-1.03677
13	1.113943	-0.06215	-0.46009	-1.12127	-1.91099	-2.69456	-3.30925	-3.44609

12	1.079181	-0.03476	0.027386	0.487474	1.608743	3.519737	6.214299	9.523547
18	1.255273	0.176091	0.210853	0.183468	-0.30401	-1.91275	-5.43249	-11.6468
14	1.146128	-0.10914	-0.28524	-0.49609	-0.67956	-0.37555	1.537199	6.969685
12	1.079181	-0.06695	0.042198	0.327433	0.823523	1.503079	1.87863	0.341431
16	1.20412	0.124939	0.191886	0.149688	-0.17775	-1.00127	-2.50435	-4.38298
26	1.414973	0.210853	0.085915	-0.10597	-0.25566	-0.07791	0.923355	3.427702
11	1.041393	-0.37358	-0.58443	-0.67035	-0.56438	-0.30872	-0.23081	-1.15416
14	1.146128	0.104735	0.478316	1.06275	1.733099	2.297476	2.606195	2.837001
20	1.30103	0.154902	0.050167	-0.42815	-1.4909	-3.224	-5.52147	-8.12767
19	1.278754	-0.02228	-0.17718	-0.22734	0.200804	1.691704	4.915702	10.43718
5	0.69897	-0.57978	-0.55751	-0.38033	-0.15298	-0.35379	-2.04549	-6.96119
9	0.954243	0.255273	0.835056	1.392563	1.772892	1.925876	2.279664	4.325157
6	0.778151	-0.17609	-0.43136	-1.26642	-2.65898	-4.43188	-6.35775	-8.63742
12	1.079181	0.30103	0.477121	0.908485	2.174905	4.833888	9.265763	15.62351
15	1.176091	0.09691	-0.20412	-0.68124	-1.58973	-3.76463	-8.59852	-17.8643
16	1.20412	0.028029	-0.06888	0.135239	0.81648	2.406206	6.170837	14.76936
21	1.322219	0.118099	0.090071	0.158952	0.023713	-0.79277	-3.19897	-9.36981
12	1.079181	-0.24304	-0.36114	-0.45121	-0.61016	-0.63387	0.158894	3.357867
12	1.079181	0	0.243038	0.604175	1.055383	1.665543	2.299416	2.140522
17	1.230449	0.151268	0.151268	-0.09177	-0.69595	-1.75133	-3.41687	-5.71629
9	0.954243	-0.27621	-0.42747	-0.57874	-0.48697	0.208974	1.960304	5.377176
21	1.322219	0.367977	0.644183	1.071657	1.650399	2.13737	1.928396	-0.03191
15	1.176091	-0.14613	-0.5141	-1.15829	-2.22995	-3.88034	-6.01771	-7.94611
26	1.414973	0.238882	0.38501	0.899115	2.057403	4.287348	8.167693	14.18541
9	0.954243	-0.46073	-0.69961	-1.08462	-1.98374	-4.04114	-8.32849	-16.4962
13	1.113943	0.159701	0.620432	1.320045	2.404668	4.388406	8.429547	16.75804
24	1.380211	0.266268	0.106567	-0.51386	-1.83391	-4.23858	-8.62698	-17.0565
14	1.146128	-0.23408	-0.50035	-0.60692	-0.09305	1.740856	5.979433	14.60642
36	1.556303	0.410174	0.644258	1.144609	1.751527	1.84458	0.103725	-5.87571
24	1.380211	-0.17609	-0.58627	-1.23052	-2.37513	-4.12666	-5.97124	-6.07496
20	1.30103	-0.07918	0.09691	0.683176	1.913699	4.288831	8.41549	14.38673
14	1.146128	-0.1549	-0.07572	-0.17263	-0.85581	-2.76951	-7.05834	-15.4738
22	1.342423	0.196295	0.351197	0.426917	0.599548	1.455355	4.22486	11.2832
20	1.30103	-0.04139	-0.23769	-0.58888	-1.0158	-1.61535	-3.0707	-7.29556

From the table the equation of the line for the model is

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \quad (16)$$

Here Y_t is taken as Y the dependent variable, while the independent variables are :

$$\Delta Yt = X_1 \quad (17)$$

$$\Delta^2 Yt = X_2 \quad (18)$$

$$\Delta^3 Yt = X_3 \quad (19)$$

$$\Delta^4 Yt = X_4 \quad (20)$$

$$\Delta^5 Yt = X_5 \quad (21)$$

$$\Delta^6 Yt = X_6 \quad (22)$$

4.4.2 Parameter Estimation

To estimate the values of the constant term β_0 and the unknown parameter $\beta_1, \beta_2, \beta_{3,\ldots,}\beta_6$. This table can be converted to matrix starting from row 9 in the table since that is where the complete set of data starts. The variable X will be taken a 40 by 7 matrix while Y will be taken a 40 by 1 matrix, and the β will be a 7 by 1 matrix.

β will be estimated by $(X^1X)^{-1} . (X^1Y)$. This is a very long process considering the number of rows involved in the matrix, since it is a long set of data it is preferable to use the statistical package for the estimation rather than matrix.

Using the Ms-Excel package, Y_t is regress on ΔYt, Δ^2Yt, Δ^3Yt, Δ^4Yt, Δ^5Yt, Δ^6Yt .

4.4.3 Prediction

Using the parameter estimates new set of values are predicted for Y which is equivalent to the differened values of the consecutive logarithm of Y. In this case study since there are a larger sets of data that can be used for this processing. The predicted values are taking in as dependent variables and are regress over newly generated sets of independent varaibles. The dependent variable(which is the predicted Y) is refer to as Zt and is shown in table 5 below

Table 5: The predicted value to be regressed

Z_t	Z_{t-1}	Z_{t-2}	Z_{t-3}	Z_{t-4}	Z_{t-5}	Z_{t-6}
0.106852						
-0.04353	0.106852					
0.01054	-0.04353	0.106852				
-0.10752	0.01054	-0.04353	0.106852			
-0.22837	-0.10752	0.01054	-0.04353	0.106852		
0.416059	-0.22837	-0.10752	0.01054	-0.04353	0.106852	
-0.05142	0.416059	-0.22837	-0.10752	0.01054	-0.04353	0.106852
-0.01464	-0.05142	0.416059	-0.22837	-0.10752	0.01054	-0.04353
0.177797	-0.01464	-0.05142	0.416059	-0.22837	-0.10752	0.01054
-0.11777	0.177797	-0.01464	-0.05142	0.416059	-0.22837	-0.10752
-0.08728	-0.11777	0.177797	-0.01464	-0.05142	0.416059	-0.22837
0.090674	-0.08728	-0.11777	0.177797	-0.01464	-0.05142	0.416059
0.184719	0.090674	-0.08728	-0.11777	0.177797	-0.01464	-0.05142
-0.37739	0.184719	0.090674	-0.08728	-0.11777	0.177797	-0.01464
0.109477	-0.37739	0.184719	0.090674	-0.08728	-0.11777	0.177797
0.153928	0.109477	-0.37739	0.184719	0.090674	-0.08728	-0.11777
-0.04392	0.153928	0.109477	-0.37739	0.184719	0.090674	-0.08728
-0.53692	-0.04392	0.153928	0.109477	-0.37739	0.184719	0.090674
0.316604	-0.53692	-0.04392	0.153928	0.109477	-0.37739	0.184719
-0.09909	0.316604	-0.53692	-0.04392	0.153928	0.109477	-0.37739
0.321896	-0.09909	0.316604	-0.53692	-0.04392	0.153928	0.109477
0.118856	0.321896	-0.09909	0.316604	-0.53692	-0.04392	0.153928
0.042414	0.118856	0.321896	-0.09909	0.316604	-0.53692	-0.04392
0.089627	0.042414	0.118856	0.321896	-0.09909	0.316604	-0.53692
-0.30132	0.089627	0.042414	0.118856	0.321896	-0.09909	0.316604
-0.02864	-0.30132	0.089627	0.042414	0.118856	0.321896	-0.09909
0.110106	-0.02864	-0.30132	0.089627	0.042414	0.118856	0.321896
-0.26109	0.110106	-0.02864	-0.30132	0.089627	0.042414	0.118856
0.370546	-0.26109	0.110106	-0.02864	-0.30132	0.089627	0.042414
-0.13617	0.370546	-0.26109	0.110106	-0.02864	-0.30132	0.089627
0.223606	-0.13617	0.370546	-0.26109	0.110106	-0.02864	-0.30132
-0.45753	0.223606	-0.13617	0.370546	-0.26109	0.110106	-0.02864
0.165921	-0.45753	0.223606	-0.13617	0.370546	-0.26109	0.110106
0.250807	0.165921	-0.45753	0.223606	-0.13617	0.370546	-0.26109
-0.25198	0.250807	0.165921	-0.45753	0.223606	-0.13617	0.370546
0.393144	-0.25198	0.250807	0.165921	-0.45753	0.223606	-0.13617
-0.20038	0.393144	-0.25198	0.250807	0.165921	-0.45753	0.223606
-0.09303	-0.20038	0.393144	-0.25198	0.250807	0.165921	-0.45753
-0.17851	-0.09303	-0.20038	0.393144	-0.25198	0.250807	0.165921
0.184531	-0.17851	-0.09303	-0.20038	0.393144	-0.25198	0.250807
-0.03451	0.184531	-0.17851	-0.09303	-0.20038	0.393144	-0.25198

-0.03451	0.184531	-0.17851	-0.09303	-0.20038	0.393144
	-0.03451	0.184531	-0.17851	-0.09303	-0.20038
		-0.03451	0.184531	-0.17851	-0.09303
			-0.03451	0.184531	-0.17851
				-0.03451	0.184531
					-0.03451

At this stage Z_t is regress on Z_{t-1} up to Z_{t-6} new set of parameters and constant are estimated this are now fit into the model equation to predict $Y_{49.}$ To predict the 49th week, The Z_{49} value has to be determined first, that is done from the equation of the model

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \quad (23)$$

$$Y_t = Z_{t,}\ X_1 = Z_{t-1,}\ X_2 = Z_{t-2,}\ X_3 = Z_{t-3,\ldots,}\ X_6 = Z_{t-6} \quad (24)$$

$$Z_{49} = \beta_0 + \beta_1 Z_{48} + \beta_2 Z_{47} + \beta_3 Z_{46} + \beta_4 Z_{45} + \beta_5 Z_{44} + \beta_6 Z_{43} \quad (25)$$

The unknown parameters and constant term were estimated during the regression and are as follows

β_0 = 0.0254, β_1=-0.75894 β_2 = -0.5281, β_3 = -0.44751, β_4=-0.49436, β_5 = -0.54157, β_6 = -0.42397

Substituting all this into equation 6, the value of Z_{49} = 0.021861

Considering table 4 above where $Y_t = \log Y_t - \log Y_{t-1}$

So, Log Y_{49} = Z_{49} + Log Y_{48}
= 0.021861 + 1.30103
= 1.322891
Log Y_{49} = 1.322891
Y_{49} = 10 ^ $^{1.322891}$
Y_{49} = 21.03251

The predicted value for the 49th week will be 21.03251 ≈ 21.

The measured value generated on the 49th week was 22 while 21 was predicted.

5. COMPARISON OF PREDICTIONS FROM BOX AND JENKINS ARIMA MODELS AND ECONOMETRIC MODELS

In previous research carried out by Somak et al where the Box and Jenkins Model was used on the two case studies, the Box-Jenkins ARIMA models predicted 66 as the value for the 10th day in the case study 1 while the econometric model predicted 75 which is closer to the measured value. The measured value of the network and system information gathering on the 10th day was 79.

While for the case study 2, The measured or collected value for the 49th week was 22. The Box-Jenkins ARIMA model predicted 19 which is close to the measured value, while the econometric model predicted 21 which is closer to the measured value than the predicted value of the Box-Jenkins ARIMA model.

6. CONCLUSION

With the results achieved, the two methods used are both suitable for predictions. Although the econometric method predicted better values to the measured values than the Box and Jenkins ARIMA method. This could have resulted from the double regression taken by the Econometric method. Although the Econometric method is more suitable for using with large datasets.

REFERENCES

[1] Arsham Hossein (Prof.) (1994), Time Series Analysis for Business forecating. http://www.mirrorservice.org/sites/home.ubalt.edu/ntsbarsh/Business-stat

[2] Barkley John F, (Nov, 1989), Introduction to Heterogeneous computing Environments,. NIST special publication

[3] Bradley Mitchell, (1999), LAN – Local Area Network, the New York Times Company.

[4] Bruce L. Bonerman and Richard T, (1993); Forecasting and Time Series an Applied Approach 3rd Edition, Encyclopedia Britanica Inc. U.S.A.

[5] Bowerman, B.L, and R.T. O'Connell, (1987), Time series forecasting, unified concepts and computer implementation, 2nd ed., Duxbury Press, Boston.

[6] Damodar N.Gujarati (2006) McGraw Hill Co. Inc. NewYork; Essentials of Econometrics

[7] David S. Walonick (1993) , An Overview of Forecasting Methods. StatPac Inc. 8609 Lyndale

[8] Eric W. Weisstein(2009), Wikipedia Autocorrelation http://en.wikipedia.org/wiki/Autocorrelation wikimedia Foundations, Inc

[9] FIPS Pub 191 (1994) Guidelines for the Analysis Local Area Network Security Federal Information Processesing Standards 191

[10] Jan Kmenta, (1971) Elements of Econometrics, Macmillian Publishing Co, Inc, New York

[11] Mario F. Triola (2001), Elementary Statistics 8th Edition, Addison – Wesley Longman Inc, USA.

[12] Mark J.E. (1997) ; Security Threats, 29th Street Press

[13] Martin James, (1989) Local Area Networks, Architectures and implementation, The Arben Group, Inc.

[14] Matt Curtin (1997) Introduction to Network Security, Kent Information Services, Inc.

[15] Michael Anissimov (2008) What is Local Area Network, wiseGeek.

[16] Micheal K. Evans, (2003) Practical Business Forecasting 2nd Edition, Black well Publishing Co. Uk.

[17] NIST/SEMATECH (2006) e-Handbook of Statistical Methods, http:// www.itl.nist.gov/div898/handbook.

[18] Paul Bourke (1998), AutoRegression Analysis (AR), Alex Sergejew and co.

[19] Peter Brockwell and Richard A Davis,(2002) Introduction to Time Series and Forecasting Springer- Verlag New-York, Inc.

[20] Rick Marcmurchie (2002) Local Area Network / Wide Area Network Security Threat Analysis, A Guide for Non-Technical Managers responsible for a corporate network

[21] Somak Bharttacharya, S.K. Gosh (2007) Security Threats Prediction in a Local Area Network Using Statistical Models. Doi.ieeecomputersociety.org/10.1109/IPDSPS.2007

[22] Webopedia 2008, Local Area Network http://en.wikipedia.org/wiki/Local area network wikimedia Foundations, Inc.

[23] Ya-lun Chou (1974) Statistical Analysis, 2nd Edition, Fresh Meadows, New York.